

# Datamining In Telecommunication Using Anova Techniques

G.Vijayasree<sup>1</sup>, Dr.Pon Periasamy<sup>2</sup>

<sup>1</sup> Ph.D.,Research Scholar, Asst.Prof.,Dept.of Comp. Science, A.V.V.M.Sri Pushpam College  
(Autonomous),Poondi,  
Thanjavur, India

<sup>2</sup> Associate Prof., Dept.of Comp. Science, Nehru Memorial College (Autonomous), Puthanampatti,  
Trichy, India.

**Abstract:** Telecom is one of the first industry to embrace Data mining technology. Generally these organizations have numerous customers. So organizations save this Data furthermore, consequently there is generation of tremendous sum of data. This Data incorporates call-detail data, client data, furthermore, system data. To handle this vast sum of Data furthermore, to find helpful Data from this Data the programmed or semi programmed strategy should be utilized as it simplifies the work. Furthermore, the Data Mining fits for this criteria. Therefore use of Data Mining innovation is must for Telecom companies. This paper portrays how Data mining furthermore, its applications helps Telecom industry in various aspects.

**Keywords:** Applications, Data, Data Mining, Telecommunication in Datamining.

## 1. Introduction

The Telecom industry generates vast sum of data. These Data incorporates call-detail Data which portrays calling conduct of that client furthermore, that Data traverse through the system that is nothing but system data. Furthermore, the clients who are subscribed to organizations portrays the client data. So these vast sum of Data was needed to be handled consequently the concept of Human based Expert frameworks came into existence. These frameworks helped to recognize the cheats furthermore, system issues. But later it came to know that this is extremely time consuming system as it includes much of learning from Human experts. Consequently this strategy was not utilized later. At that time the Data mining was new. Data Mining is search for the relationships furthermore, global designs that exist in vast database but are hidden vast amounts of data. This relationship represents valuable learning about the database, furthermore, the objects in the database. So Telecom organizations started to use this technology. They were the first to embrace Data mining technology. But Telecom Data pose extremely intriguing issues for Data mining. The first issue is Telecom Data might contain billions of records furthermore, amongst largest in world. The second one often raw Data is not always suitable for Data mining. Furthermore, the last issue is concerned with the real time execution of Data mining applications such as misrepresentation identification which requires learned protocol, rules to be connected in real-time. If we bargain with these issues then the efficiency of Data mining will definitely increase. But if we just consider its usefulness then we come to know that it is a fundamental tool for the industry.

## 2. Data Mining

Data Mining implies extricating learning hidden in vast volumes of data. It is a part of learning disco extremely process that offers new way to look at data. Simply, identifying potentially helpful furthermore, understandable data. It is the progressed investigation step of the Learning Disco extremely in database. The real Data mining task is the programmed or semi-programmed investigation of vast quantities of Data to extricate previously unknown intriguing designs such as bunches of Data records (group analysis), unusual records (anomaly detection) furthermore, dependencies (association rule mining). This usually includes utilizing database methods such as spatial indexes. These designs can then be seen as a kind of summary of the input data, furthermore, might be utilized in further investigation or, for example, in machine learning furthermore, predictive analytics. In straightforward terms it employs machine learning, statistical furthermore, visualization methods to find furthermore, present the learning in a form that is effortlessly comprehensible to humans. Typically in Telecom industry the Data mining process includes Client segmentation, Profiling, Data Arrangement furthermore, Clustering. These assignments are discussed in this paper.

## 3. Need of Data Mining In Telecom

As Data is the base of Telecom so Data mining is utilized to perform operations on Data furthermore, get the wanted results. Some of the reasons to use Data mining are as follows:

### ***3.1 To recognize cheats***

Misrepresentation is extremely serious issue for Telecom companies, resulting in billions of dollars of lost

### ***3.2 To retain clients***

By applying Data mining apparent employments we can study furthermore, research on the customers' database which will help to know how to bargain with our clients furthermore, how to satisfy them. In this way we can maintain good relations with the old clients as well as get new ones.

### ***3.3 To know the Clients***

By studying the conduct of the clients we can better know them, what is their calling period, their likes, tastes, attitudes etc. This will help to approach them in terms of marketing activity.

### ***3.4 Products furthermore, administrations which yield highest sum of benefit***

As it is extremely vast database, so it contains exchanges made by customers. That exchanges maybe of purchased product, services. The study of these exchanges can give us statistics that which of my administrations furthermore, products are popular that yield highest sum of profit.

### ***3.5 Factors that impact clients to call more at certain times***

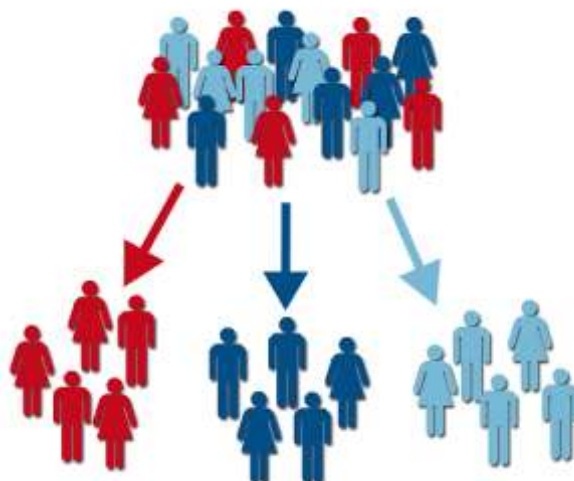
By studying conduct of customers, we can come to know what are the factors that impact clients to call more at certain times.

## **4. Client Division Furthermore, Profiling**

Client Division is the process of dividing clients into the homogeneous bunches according to their normal attributes. These characteristics include their likes, habits, actions. Furthermore, Client profiling implies to portray the clients by their characteristics like age, gender, economic conditions, income, culture etc. This Client division furthermore, profiling helps the Telecom organizations to decide which marketing activities should be taken for specific portion furthermore, allocating the resources. Client division process requires the learning of an expert. First of all the divisions are defined in advance then clients are divided over these segmentations. Furthermore, then the profile for extremely client is determined with the help of client data. Now, to find the relation between client profile furthermore, sections the SVM method can be used. SVM is a Data Mining method which stands for Support Vector Machine [2]. SVM estimates portion of a client by his profile Data (age, gender, etc). So if the portion is estimated based on client profile then we can effortlessly determine use conduct of that specific client profile.

In today's competition world to compete with other Telecom companies, it is extremely much essential to know about the clients furthermore, their needs. Furthermore, Consequently Client Division furthermore, Profiling is important. Overall the goal of doing all this is to predict the conduct of client utilizing all the Data we have. The profiling is the process done after the segmentation.

### **4.1 Client Segmentation**



## Figure 4.1 Client Segmentation

In straightforward words, it is act of partitioning the client bunches as per their comparable characteristics. Client division helps us to know about the conduct of the client like what is his usage, calling time, frequency of calling. This helps organizations to approach furthermore, to have communication with clients so that their needs can be understood furthermore, marketing in this way can be done effectively. Consequently more assets can be channeled effectively. But while doing the division some difficulties arises like-

### A.Non Applicable Data

Although company have vast sum of Data but if it is not applicable to need then it is useless. Consequently just the applicable Data should be chosen.

### B.Not Updating Data

Here we should understand furthermore, that division is continuous process furthermore, it should be redesigned frequently whenever the new client gets added.

### C.Over Division

While defining divisions care should taken that specific portion should not be too small that it will be troublesome to treat separate segments.

## 4.2 Client Profiling

Client Profile is built on the basis of his individual Data such as age, gender, location.

It is done so as to reach the clients furthermore, serve them better. Too as per the needs like to offer/manufacture new item the applicable profiles are selected. The essential characteristics of individual Data includes- • Age furthermore, Gender-Age furthermore, gender of customers.

- Location- The geographic, national region of client .
- Economic conditions- Salary of a client or his family, Payment mode.
- Attitude- The attitude of client towards utilizing the item
- Lifestyle- Lifestyle of client as per he employments the service
- Knowledge- How marketing can be done for specific clients as per their learning

## 5. Types of Telecom Data

Before performing the Data mining process it is fundamental to understand furthermore, the Data i.e. what sort of Data it is. So basically there are main three types of Data viz. Call-Detail data, System Data furthermore, Client data.

### 5.1 Call-Detail Data

Call-detail Data is a Data which contains details about when a call is placed, duration of call i.e. everything about the call. Everyday vast numbers of calls are created so when call takes place on the system then a detailed Data about call is created furthermore, is saved. That is nothing but call detail data. This Data incorporates location, call duration, call time, calls originated/received, etc. That is it contains sufficient Data to portray essential characteristics of each call. The Data about each call is spared in position of sequence of records in database. Then the summarization of all these records is done furthermore, real Data mining process is connected to extricate helpful learning as per our needs.

However the call description includes:

- normal call duration
- % no-answer calls
- % calls to/from a different region code

- % of weekday calls (Monday – Friday)
- % of daytime calls (9am – 5pm)
- normal # calls received per day
- normal # calls originated per day
- # unique region codes called amid P

From this, the Data graphs, histograms furthermore, tables are drawn. Furthermore, then these are analyzed furthermore, utilized for knowing current status furthermore, future predictions.

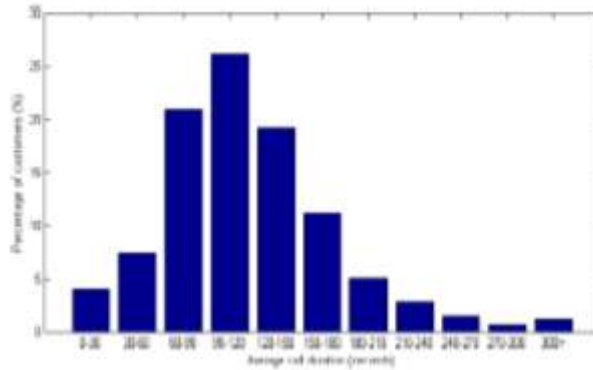


Figure 5.1 Call Duration

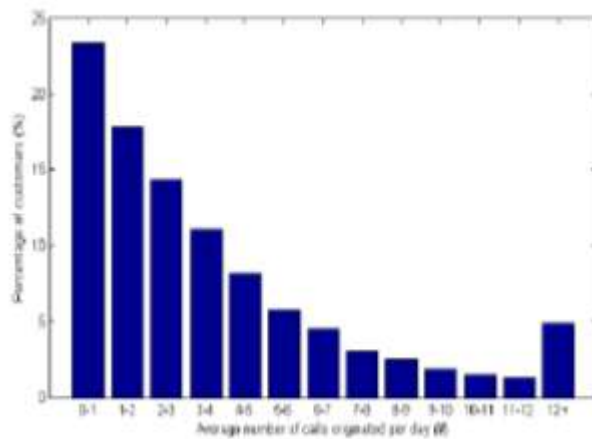


Figure 5.1 (a) Originated calls

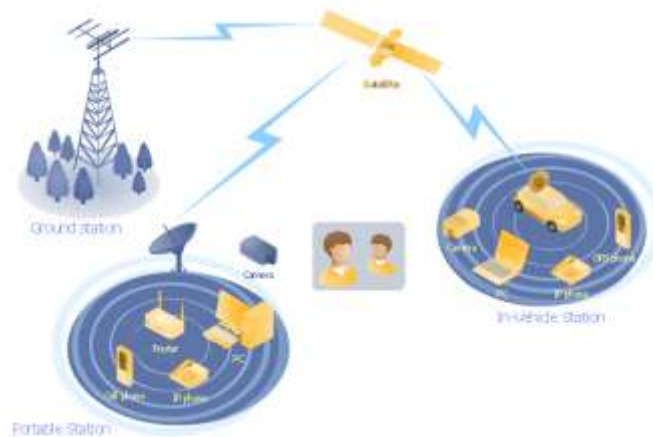
For example, the above histograms appears the relationship between clients furthermore, calls. Fig. V.(a) appears call duration of clients i.e. 25% of clients call for about 90-120 seconds. Fig. V.(b) gives Data about normal number of calls created by customers.

The call detail records are too utilized to distinguish between Businesses furthermore, Residential calls. For example, the calls made amid the time period from 9-4/5 pm demonstrates it is business customer. Furthermore, the calls which take place mostly at evening time furthermore, on weekdays demonstrates a residential customer.

## 5.2 System Data

The Telecom system is extremely complex in nature. It consists of numerous elements furthermore, components. Sometimes a blunder might occur in these equipment's while being in network. So they create blunder furthermore, status messages. Furthermore, these messages should be stored in database so as to support system management functions.

Suppose if specific component is down for some time or it is overloaded furthermore, if status messages indicating these problems are stored then proper action can be taken to resolve them by analysis. But extremely time manual investigation is difficult. Hence, Data mining innovation is utilized to recognize furthermore, resolve the system flaws by extricating Data from system data.



**Figure 5.2 System Architecture**

### 5.3 Client Data

Table 5.3 Client data

Age	<25 25.2%	25-40 32.5%	40-55 36.9%	>55 26.4%
Gender	Male 58.2%	Female 31.8%		
Telephone type	Simple 36.5%	Basic 38.7%	Advanced 35.8%	
Type of Subscription	Simple 36.9%	Advanced 33.0%	Expanded 30.1%	
Company size	Small 31.5%	Intermediate 34.3%	Big 34.2%	
Living area	(big city) 42.2%	Small 58.0%		

The Telecom organizations have gigantic sum of customers. So Data about each of them needs to be maintained. Client Data incorporates details about their individual Data like age, gender, salary furthermore, phone type, subscription type. Maintaining this client Data essential so as to know profiles of customers. In the above chart certain parameters are given which demonstrates current status of them like there are total 29.5% clients of age 25-40 furthermore, too there are 21.2% of clients whose age is less than 25. The 38% clients use essential phone sort furthermore, 27.8% use advanced.

## 6. Data Arrangement Furthermore, Clustering

These are the assignments of Data Mining. They are discussed below:

### A. Data Arrangement

Before utilizing real Data mining process the Data should be prepared in the wanted format. So Data Arrangement includes numerous tasks. They are

- Find helpful Data furthermore, repair inconsistency between these Data formats
- Uproot the spelling mistakes, furthermore, check for syntax errors
- Uproot fundamental Data fields
- Normalize the tables
- Check blank Data fields

## B. Grouping

Grouping implies grouping of comparable things. So here we can see how Grouping is done in Data mining.

### Group Analysis:

A group implies collection of objects which are comparable between them furthermore, incomparable to the objects belonging to other objects. The group investigation is done to organize objects into their specific bunches as per the likenesses between them.

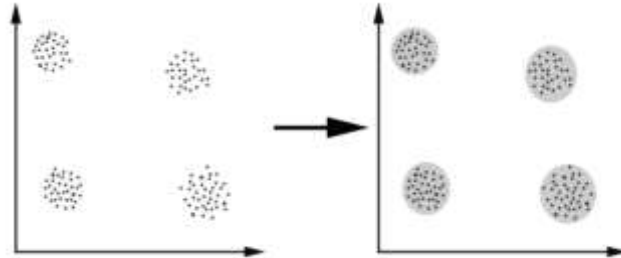


Figure. 6.1 Clustering

In this case in first three clusters, likenesses are found furthermore, then they are grouped accordingly. Equalize the length of your columns on the last page. If you are using *Word*, proceed as follows: Insert/Break/Continuous.

## 7. Data Mining Applications

### 7.1 Misrepresentation Identification



Figure. 7.1 Misrepresentation Identification

Misrepresentation is a dangerous issue for the Telecom industry. It results in gigantic loss to them. The normal strategy to recognize misrepresentation is to manufacture customer's profile of calling conduct which contains his previous calling pattern. Then his recent action is compared against this behavior. So if any suspicious action is seen then that misrepresentation is caught. Thus, the Data mining application depends on deviation detection. The call-detail records are summarized to get calling behavior. If the call detail summaries are redesigned in ongoing then the misrepresentation can be detected soon after it occurs.

### 7.2 Client Profiling



Fig. 7.2 Marketing Approach

Marketing is an essential thing for the Telecom companies. The organizations maintains gigantic sum of Data about their customers. This Data incorporates his individual Data like age, gender, lifestyle, learning etc. which might be stored in Data mart. Then the Data mining apparent employments are connected furthermore, divisions is done on this vast sum of data. Furthermore, that is utilized to manufacture customers' profile furthermore, then it is helps to manufacture marketing strategies, planning for the future decisions, execution measurement, result tracking, etc. Consequently organizations come to know who are my customers?, what is their gender?, what are their needs?, furthermore, how I can approach furthermore, offer new administrations to them? Therefore Client profiling is an essential aspect.

### 7.3 System Fault Isolation

As mentioned earlier Telecom networks are of extremely complex configuration furthermore, they consist of numerous interconnected components. These system parts create status as well as alarm messages extremely time. So in order to recognize the system flaws the alarms should be analyzed automatically. So the Data mining helps to naturally analyze furthermore, recognize the flaws so that they can be resolved.

### 7.4 Tools used

SPSS is a widely used program for statistical analysis in social science. It is also used by market researchers, health researchers, survey companies, government, education researchers, organizations, data miners, and others. We have used this software for analyzing the data. The version which we used is IBM SPSS 19.0.

#### A. Differences between Communication Profiles

A one-way between- groups study of modifications (ANOVA) was showed to discover whether the experimental differences in requisite functions for the cluster were significant. This analysis was carried out separately for each task. There was a statistically significance difference at the  $p < .05$  level in average requisite function scores for all five communication profiles for the decision making and production tasks. These significant differences are summarized in table 7.1.

Table 7.1 Statistical significance level of between- groups ANOVA Exploring cluster differences

Requisite Function	Significance level	
<b>NASA TASK</b>		
Orientation	F (4, 198)=50.48	$p < .001$
Problem Definition	F (4, 198)=38.45	$p < .001$
Criteria Development	F (4, 198)= 2.53	$p < .05$
Solution Development	F (4, 198)= 134.35	$p < .001$
Solution Evaluation	F (4, 198)=5.13	$p < .001$
<b>ORIGAMI TASK</b>		

Orientation	F (4, 198)=58.39	p<.001
Problem Definition	F (4, 198)=40.59	p<.001
Criteria Development	F (4, 198)= 2.71	p<.05
Solution Development	F (4, 198)= 110.61	p<.001
Solution Evaluation	F (4, 198)=5.13	p<.001

p<.05, p<.01, p<.001.

## 8. Conclusion

The telecom industry is the early adopter of Data mining technology. It reduced much of human based analysis. It has extremely helpful applications. It helps to recognize frauds, to know the clients furthermore, serve them better as per their needs. Furthermore, apart from client side, to come to know how we can yield more profit. So overall it is extremely much essential in fact must for Telecom companies.

## References

- [1] Jung-Kyu Choi; Dept. of Healthcare Eng., Chonbuk Nat. Univ., Jeonju, South Korea; Chan Il Yoo; Kyung-Ah Kim; Yonggwon Won, "Study on Datamining Technique for Foot Disease Prediction", Published in: IT Convergence and Security (ICITCS), 2014 International Conference on Date of Conference: 28-30 Oct. 2014. Page(s): 1 – 4
- [2] Emanuel, A.W.R.; Fac. of Inf. Technol., Maranatha Christian Univ., Bandung, Indonesia; Wardoyo, R.; Istiyanto, J.E.; Mustofa, K., "Success factors of OSS projects from sourceforge using Datamining Association Rule", Published in: Distributed Framework and Applications (DFmA), 2010 International Conference on Date of Conference: 2-3 Aug. 2010 Page(s): 1 – 8.
- [3] Gulski, E.; Delft Univ. of Technol., Netherlands; Quak, B.; Wester, F.J.; de Vries, F., "Application of datamining techniques for power cable diagnosis", Published in:
- [4] Properties and Applications of Dielectric Materials, 2003. Proceedings of the 7th International Conference on (Volume:3 ) Date of Conference: 1-5 June 2003 Page(s): 986 - 989 vol.3.
- [5] Erraguntla, M.; Ramachandran, S.; Chang-Nien Wu; Mayer, R.J., "Avian Influenza Datamining Using Environment, Epidemiology, and Etiology Surveillance and Analysis Toolkit (E3SAT), Published in: System Sciences (HICSS), 2010 43rd Hawaii International Conference on Date of Conference: 5-8 Jan. 2010 Page(s): 1 – 7.
- [6] Jiang, T.; Xu, Y.; Zhang, M., "Datamining: Discovering Information from Bio-Data", Publisher: MIT Press Edition: 1 Pages: 317 – 341.
- [7] Llorente, R.; Univ. Politec. de Valencia, Valencia, Spain; Morant, M., "Accurate knowledge evaluation by deep datamining in Telecommunication Engineering studies", Published in: EAEEIE Annual Conference, 2009 Date of Conference: 22-24 June 2009 Page(s):1 – 6.
- [8] Kantardzic, M., "Appendix B: DataMining Applications", Publisher: Wiley-IEEE Press Edition: 1 Pages: 327 – 338.
- [9] Singh, M.P.; Comput. Network & Internet Eng., Center For Dev. Of Adv. Comput., Bangalore, India; Subramanian, N.; Rajamenakshi, "Visualization of flow data based on clustering technique for identifying network anomalies", Published in: Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on (Volume:2 ) Date of Conference: 4-6 Oct. 2009 Page(s): 973 – 978.
- [10] Min Qin; Southern California Univ., Los Angeles, CA, USA; Kai Hwang, "Frequent episode rules for Internet anomaly detection", Published in: Network Computing and Applications, 2004. (NCA 2004). Proceedings. Third IEEE International Symposium on Date of Conference: 30 Aug.-1 Sept. 2004 Page(s): 161 – 168.
- [11] Thongraweewong, K., "State telecommunication surveillance: A comparative study of the US and Thai telecommunication privacy laws", Published in: Digital Information and Communication Technology and it's Applications (DICTAP), 2014 Fourth International Conference on Date of Conference: 6-8 May 2014 Page(s): 347 – 352.
- [12] Evert, G.P.; Dept. of Eng. & Technol. Manage., Univ. of Pretoria, Pretoria, South Africa; Erasmus, L.D., "The adoption and effects of international telecommunication trends in South Africa: Technology and market perspectives", Published in: Management of Engineering & Technology (PICMET), 2014 Portland International Conference on Date of Conference: 27-31 July 2014 Page(s): 2039 – 2046.